Student Stress Level Predictors

By: Daniel Meier, Joseph Choi, Jenny Albrecht and Serge Nane

Why this topic?

- Stress is a universal issue for all, especially for students.
- As current students, we were interested in exploring this topic for our own personal interests
- Determining factors that have the highest relation to stress, which allows us to take actionable measures to mitigate stress

Project Data

- Goal of Project:
  - Create a predictive model to identify both a student's level of stress and the greatest predictors of stress level
- Dataset origin:
  - Kaggle (Student Stress Factors: A Comprehensive Analysis (kaggle.com))
- About the data:
  - Stress factors broken into 5 categories - Psychological, Physiological, Environmental, Academic, and Social
  - 1100 data points, 21 features total

Data Wrangling

Data Wrangling Steps

- Loaded in our dataset and performed some initial analysis
- Cleaned data that showed as categorical but needed to be numerical
  - Anxiety_level, self_esteem and depression
- No missing values
- Most features demonstrated a balanced distribution (compared mean of each feature to range of that feature)

Data Science

Exploratory Data Analysis (EDA)

- Histograms
  - Assessing the distribution of the data (skewed vs. normal)
- Correlation Matrices
  - All features had some correlation with stress
  - Determined scaling needed to occur due to larger distribution of some features (ex. self_esteem ranges 0-30, where social_support ranges 0-5)

Model Selection

- Random Forest Classifier and K-Nearest Neighbors (KNN) were selected
- These models were chosen for the following reasons:

1) Data has standardized scales and categorical rating

2) Random Forest excels at capturing non-linear relationships between input and target

3) KNN performs well with numerical data

Model Iterations

- We have chosen to conduct 5 iterations of each model
    1) All features, unscaled data
    2) Use Grid Search to tune model, all features, unscaled data
    3) All features, scaled data
    4) Use Grid Search to tune model, all features, scaled data
    5) Adjust training/testing ratio (70/30, 80/20, 90/10)

Model Performance

Best Performing Model Iteration:

Random Forest Classifier

Training Accuracy: 89.2%

Testing Accuracy: 86.4%

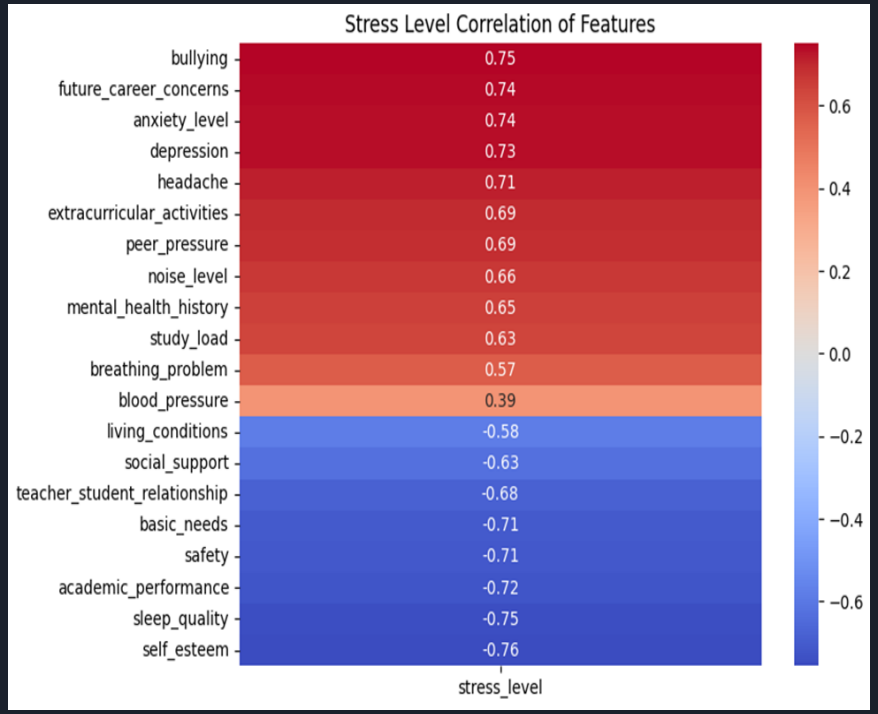
Data Visualization

Distribution of Features

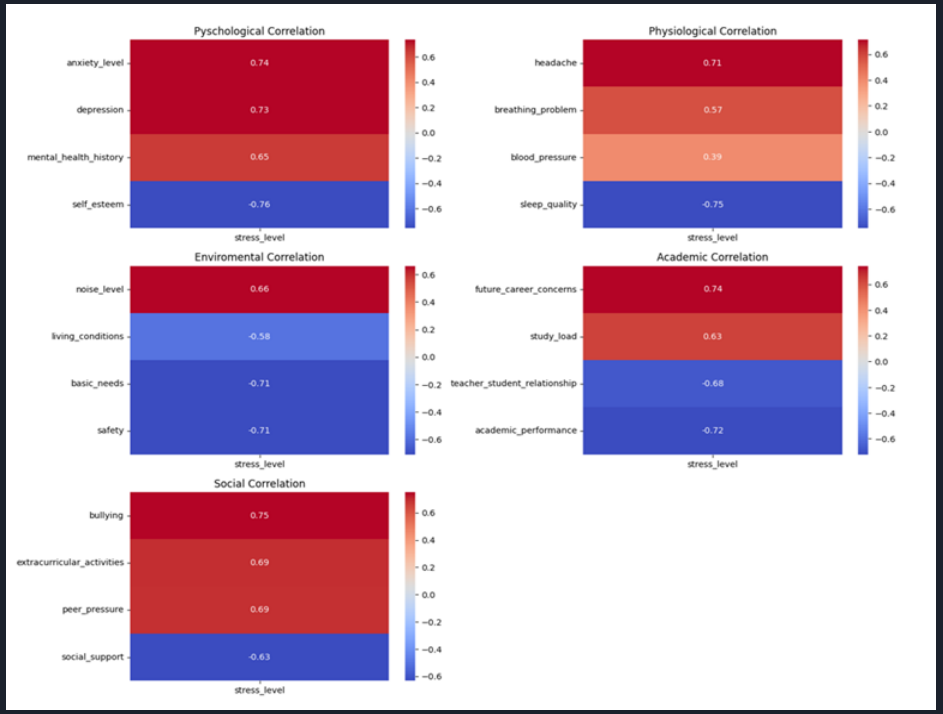
Features	Category	Remark
Anxiety Level	Evenly Distributed	Level is even throughout students
Self-Esteem	Skewed	Most students rated high
Mental Health History	Evenly Distributed	Level is even throughout students
Depression	Evenly Distributed	Level is even throughout students
Headache	Skewed	Most students rated either low or high
Blood Pressure	Skewed	Most students rated high
Sleep Quality	Skewed	Most students rated either high or low
Breathing Problem	Skewed	Most students rated either high or medium
Noise Level	Evenly Distributed	Level is even throughout students
Living Conditions	Evenly Distributed	Level is even throughout students
Safety	Skewed	Most students rated either high or medium
Basic Needs	Skewed	Most students rated either high or medium
Academic Performance	Skewed	Most students rated either high or medium
Study Load	Evenly Distributed	Level is even throughout students
Teacher-Student Relationship	Skewed	Most students rated either high or medium
Future Career Concerns	Skewed	Most students rated either high or medium
Social Support	Skewed	Most students rated either high or low
Peer Pressure	Skewed	Most students rated either high or medium
Extracurricular Activities	Skewed	Most students rated either high or medium
Bullying	Skewed	Most students rated either high or low
Stress Level	Evenly Distributed	Level is even throughout students

Correlation Matrices

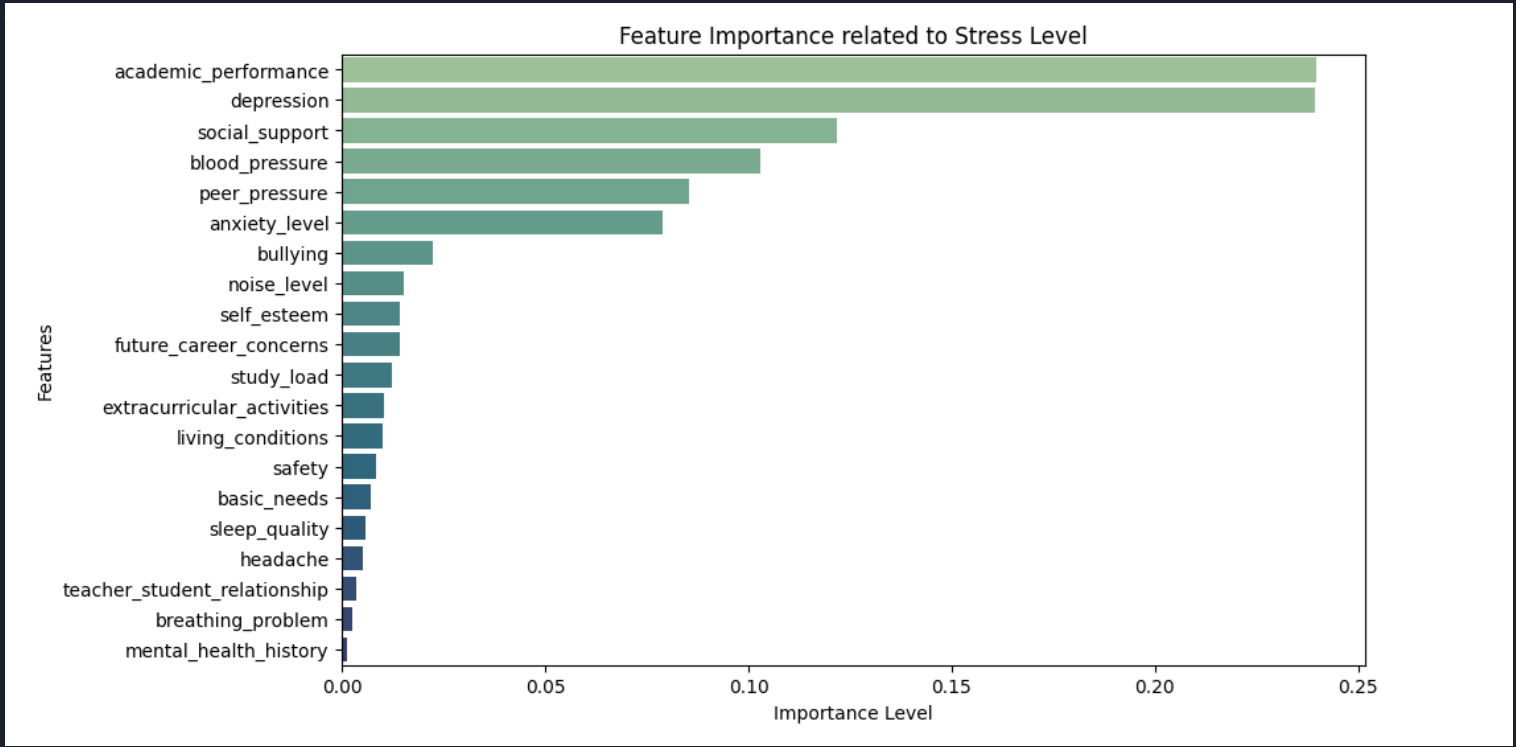
Individual



Categorized



Feature Importance



Recommendations/ Conclusion

Recommendations:

- Educate teachers, parents, students about the highest predictors of stress
- Encourage action to be taken to either increase items that decrease stress (Self-Esteem, Sleep Quality, Academic Performance) while decreasing items that increas stress (Bullying, Future Career Concerns, Anxiety Levels)

Conclusion:

- Model was very successful in its predictive power to identify stress levels in students
- Implementing this model can help promote student mental health and foster a more supportive environment for students

References/Resources

1. How stress-related factors affect mental wellbeing of university students A cross-sectional study to explore the associations between stressors, perceived stress, and mental wellbeing - PMC (nih.gov): Discusses a study that was done to understand the associations between stressors, perceived stress, and mental well-being of students attending universities.
2. Full article: The impact of stress on students in secondary school and higher education (tandfonline.com): Analyzes the impact of academic-related stress on students' mental health and how it affects academic performance.
3. Stress and Quality of Life Among University Students: A Systematic Literature Review - ScienceDirect: Provides information relating to the relationship between stress and quality of life among university students.
4. (PDF) Stress among students: An emerging issue (researchgate.net): Highlights the negative and positive impacts on the health and academic performance of students due to stress.
5. Full article: Academic stress as a predictor of mental health in university students (tandfonline.com): Dives into the relationship between academic stress and mental health among 1,265 university students
6. The Effects of Stress on College Students & Ways to Overcome it (bau.edu): Talks about the concept of academic stress and the psychological effects it has on college students.
7. Stress modelling and prediction in presence of scarce data - ScienceDirect: Discusses approaches for predicting stress levels using techniques like semi-supervised learning and ensemble methods.
8. Prediction of stress levels in the workplace using surrounding stress - ScienceDirect: Explores the prediction of stress levels by analyzing stress data. Achieved an 80% F-score using surrounding stress data.
9. Stress prediction using micro-EMA and machine learning during COVID-19 social isolation - ScienceDirect: Introduces a stress prediction system using micro-EMA historical data to forecast stress. Achieved absolute error of 4.26 and an accuracy of 81% in stress prediction.
10. Sensors | Free Full-Text | Stress Monitoring Using Machine Learning, IoT and Wearable Sensors (mdpi.com): Discusses a machine learning-based system that monitors stress by analyzing body temperature, sweat, and motion rate. Achieved an accuracy rate of 99.5%.